PHYSICAL REVIEW X **6,** 021028 (2016)

Open-System Quantum Annealing in Mean-Field Models with Exponential Degeneracy[*]

Kostyantyn Kechedzhi[1] and Vadim N. Smelyanskiy[2]

[1]*QuAIL, NASA Ames Research Center, Mail Stop 269-3, Moffett Field, California 94035 USRA,*
NASA Ames Research Center, Moffett Field, California 94035, USA
[2]*Google, 150 Main Street, Venice Beach, California 90291, USA*

(Received 9 September 2015; revised manuscript received 18 February 2016; published 31 May 2016)

Real-life quantum computers are inevitably affected by intrinsic noise resulting in dissipative nonunitary dynamics realized by these devices. We consider an open-system quantum annealing algorithm optimized for such a realistic analog quantum device which takes advantage of noise-induced thermalization and relies on incoherent quantum tunneling at finite temperature. We theoretically analyze the performance of this algorithm considering a p-spin model that allows for a mean-field quasiclassical solution and, at the same time, demonstrates the first-order phase transition and exponential degeneracy of states, typical characteristics of spin glasses. We demonstrate that finite-temperature effects introduced by the noise are particularly important for the dynamics in the presence of the exponential degeneracy of metastable states. We determine the optimal regime of the open-system quantum annealing algorithm for this model and find that it can outperform simulated annealing in a range of parameters. Large-scale multiqubit quantum tunneling is instrumental for the quantum speedup in this model, which is possible because of the unusual nonmonotonous temperature dependence of the quantum-tunneling action in this model, where the most efficient transition rate corresponds to zero temperature. This model calculation is the first analytically tractable example where open-system quantum annealing algorithm outperforms simulated annealing, which can, in principle, be realized using an analog quantum computer.

DOI: 10.1103/PhysRevX.6.021028

Subject Areas: Condensed Matter Physics,
Quantum Physics,
Quantum Information

I. INTRODUCTION

Quantum computing hardware is affected by a substantial level of intrinsic noise and therefore naturally realizes dissipative quantum dynamics [1,2]. Optimization algorithms, where a configuration of a binary string x minimizing a given (energy or cost) function $f(x)$ is sought for, naturally extract a computational advantage from the irreversible dissipative dynamics and could therefore be readily implemented on a number of existing hardware platforms [3,4]. More specifically, quantum annealing [5] (QA) is a quantum analog of the widely applied classical simulated annealing algorithm [6] (SA), a heuristic solver of NP-hard (non-deterministic polynomial-time hard [7]) optimization problems [5,8–11], with quantum fluctuations playing the role analogous to thermal fluctuations in simulated annealing. NP-hard optimization problems, such as finding a ground-state spin configuration of a spin glass, are often characterized by an energy landscape with a large number of local minima separated by extensive energy barriers. Dissipative dynamics realized by the open-system quantum annealing provides an efficient mechanism for thermalization within domains of attraction of local minima. For an efficient search of the configuration space, the barriers separating different domains of attraction have to be overcome, which may proceed via thermal excitation or a quantum-tunneling process. The performance of the open-system quantum annealing algorithm is therefore characterized by a set of relaxation rates associated with such processes, as opposed to, for example, the spectral gaps, as is the case for an adiabatic quantum algorithm [10,12,13].

The longest relaxation times correspond to the often exponentially slow transitions between local minima separated by extensive potential barriers. Unitary dynamics of a pair of such states corresponds to the switching rate of the order of the matrix element Δ, which in the presence of an extensive barrier may scale exponentially with the system size $\Delta \propto \exp(-\mathrm{const} \times N)$ (here, N is the number of qubits in the system). However, fast dissipative relaxation within a domain of attraction of a local minima due to the hardware noise introduces a lifetime or level width W. This fast local relaxation strongly suppresses the coherent superposition

---

[*] The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon.

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

of the states localized in different local minima when $W \gg \Delta$. Nevertheless, the incoherent quantum tunneling is possible in the presence of such strong dissipation, where the transition rate is described by the Fermi golden-rule-type expression $\propto \Delta^2 \ll \Delta$. This is the regime likely realized in a large-scale quantum annealer [3,14]. It is an open question whether such incoherent extensive quantum tunneling may provide a more efficient mechanism for searching the configuration space as compared to classical simulated annealing relying on thermal excitation. Initial numerical analysis of the two-dimensional (2D) spin-glass problem [15] suggested a superior scaling of the outcome of finite-temperature quantum annealing compared to simulated annealing [11]. However, this observation turned out to be an artifact of the numerical discretization scheme and therefore cannot be reproduced using analog hardware [17].

In this paper, we provide an analytically tractable example where the speedup originates from incoherent quantum tunneling. We consider a system of Ising spins interacting each with each other with a p-body interaction of equal strength, a model often referred to as a p-spin model. This model allows a quasiclassical Wentzel-Kramers-Brillouin (WKB) description [18,19], where the expansion is performed in $1/N$ rather than the more usual $\hbar$, and, at the same time, it demonstrates key features characteristic of a range of complex (NP-hard) optimization problems, such as the first-order phase transition (for $p \geq 3$) and an exponentially small gap between the ground and excited states [20]. Crucially, the metastable state realized in this model is characterized by an exponential degeneracy, whereas the ground state is unique. Such entropic imbalance is, in fact, typical for low-energy states in the spin-glass phase, and it strongly affects the low-temperature dynamics of the system in both quantum and classical cases. The effect of entropic imbalance is the main focus of our analysis in this paper.

We show that the scaling of the optimal QA computation time (allowing for repeated runs of the algorithm) is determined by the quantum-tunneling amplitude at a single point in the algorithm, the so-called freezing point, after which quantum (or thermal) fluctuations are relatively weak and the transitions over or through the barrier are no longer likely. We find that, because of the effect of the entropy associated with the metastable state, the optimal quantum-tunneling rate is achieved at vanishing temperature; i.e., raising the temperature may reduce the quantum-tunneling rate. This is in contrast to the usual intuition about a quantum-mechanical particle trapped in a nondegenerate metastable potential well, where the escape rate monotonously increases with temperature. The optimal QA regime corresponds to the fastest tunneling rate and therefore vanishing temperature as well. Comparing the optimal computation time of QA obtained in this regime with that of SA for a range of the potential barrier shapes, we find that QA could outperform SA under certain circumstances, thus providing a polynomial (rather than exponential) speedup.

The physical mechanism of the quantum speedup that we find is distinct from the usual intuition of quantum fluctuations overcoming thin and tall barriers more efficiently. It is a generic mechanism, and we expect it to manifest in more complex quantum models where quantum fluctuations are introduced by transverse spin-spin or multispin interaction terms [21–23] in addition to the transverse field. We emphasize that the speedup mechanism we find is not limited to the models that can be efficiently simulated with quantum Monte Carlo dynamics on classical computers [24,25] (such as the transverse-field Ising model). The regime considered here could, in principle, be reproduced using an analog quantum annealer.

Before outlining the formal calculation, we discuss the qualitative picture of the effect of the entropy imbalance between the ground and metastable states on the efficiency of simulated and quantum annealing. In the presence of the entropy imbalance, SA computation time scales exponentially with the system size N. This can be understood intuitively by considering the performance of SA applied to a model demonstrating a first-order phase transition into a state characterized by an order parameter, assuming a ferromagnet for simplicity. In simulated annealing, the system is initialized at infinite temperature, or equal occupation of all classical spin states, and then the temperature is gradually lowered to zero. The simulated spin dynamics is chosen to satisfy the detailed balance condition such that it samples the thermal distribution at a given (instant) temperature. The initial state is a paramagnet, and therefore the solution, the ground-state spin configuration at zero temperature, is expected to have high statistical weight only at low enough temperatures below the ferromagnetic phase transition. The exponential degeneracy of the metastable state corresponds to the entropy linear in the system size N, which significantly lowers the transition temperature (see Fig. 1).

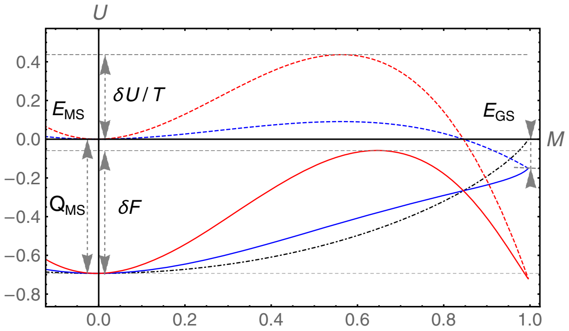

FIG. 1. Thermodynamic functions of a classical model characterized by the order parameter M above (blue lines) and below (red lines) the phase transition temperature. Dashed lines correspond to potential energy, and solid lines correspond to free energy that includes the entropy, shown separately by a dash-dotted black line. The difference between dashed and solid lines shows the contribution from entropy. δF is the free energy difference that has to be overcome by classical Monte Carlo dynamics at temperatures below the transition $T < T_c$.

This can be understood intuitively from the following argument. We assume a mean-field case in which energies of the metastable and ground states, as well as the barrier separating them, scale linearly with the system size NE_{MS}, NE_{GS} and NU, respectively. We find from equating the free energies $NQ_{\mathrm{MS}} - NE_{\mathrm{MS}}/T \approx -NE_{\mathrm{GS}}/T$, where the entropy imbalance is given by NQ_{MS}, that the transition occurs at $T_c \sim [(E_{\mathrm{MS}} - E_{\mathrm{GS}})/(Q_{\mathrm{MS}})] \sim O(1)$. Furthermore, high statistical weight of the ground state is achieved only after equilibration at the low temperature below the phase transition $T \lesssim T_c \sim O(1)$. In the presence of the extensive barrier NU, the relaxation towards the thermal distribution described by the classical Kramers escape rate $\sim \exp(-NU/T_c)$ is exponentially slow. Moreover, the entropy gradient along the over-the-barrier escape trajectory gives rise to an additional entropic factor $\sim \exp[N(Q_{\mathrm{MS}} - Q_T)]$, where Q_T is the entropy corresponding to the maximum of the free energy. Here, Q_T appears as an exponential contribution to the prefactor of the Kramers rate [26]. In this paper, we call Q_T the "entropic barrier." The SA computation time allowing for such relaxation to occur is at least as long as the relaxation time $\tau_u \sim \exp[NUQ_{\mathrm{MS}}/(E_{\mathrm{MS}} - E_{\mathrm{GS}}) + N(Q_{\mathrm{MS}} - Q_T)]$. In fact, a more careful analysis (see Appendix A) shows that the optimal SA computation time in the presence of the extensive entropy imbalance is given by the smallest of either τ_u or the exhaustive search time $\tau_{\mathrm{es}} \sim 2^N$. At the same time, the quantum-tunneling amplitude saturates as $T \to 0$ and may be more efficient than over the barrier escape, suggesting that quantum annealing could be more efficient than simulated annealing. Moreover, we see below that the transverse field introduced in QA lifts the degeneracy of the metastable state, and in this way, QA avoids the entropic barrier completely. Note, however, that the quantum-tunneling rate in this mean-field model also scales exponentially with N. Therefore, the performance (computation time) of SA and QA is characterized by the numerical scaling factors in the exponent (in front of N), which have to be carefully compared. The result is not obvious *a priori* since here we are comparing different microscopic mechanisms: the quantum dynamics constrained by conservation laws with the classical thermal excitation process constrained by the entropy imbalance and the low temperature.

The remainder of the paper is organized as follows. In Sec. II, we introduce the p-spin model and its WKB analysis describing the evolution of the potential energy and the transition rates in the course of the quantum annealing algorithm. In Sec. III, we discuss the dynamics of the model in the course of the quantum annealing and identify the freezing point and its optimal position in the course of the algorithm. We conclude with a discussion of the results in Sec. IV.

II. THE MODEL

We consider N Ising spins-1/2 on a fully connected graph (i.e., each spin interacts with each other spin) with uniform strength of spin-spin and multispin interactions, subject to a uniform field. The uniform interactions give rise to a highly symmetric Hilbert space such that the unitary dynamics of the system is fully described in terms of the total spin projection operators $\hat{S}^\alpha = \sum_{i=1}^N \hat{\sigma}_i^\alpha$, where $\alpha = x, y, z$ and $\hat{\sigma}^\alpha$ is a set of spin-1/2 operators. In other words, the Hamiltonian is defined as

$$\mathcal{H} = sNf\left(\frac{2}{N}\hat{S}^z\right) - (1-s)\hat{S}^x. \tag{1}$$

Here, the second term describes the uniform transverse field, and the first term is a potential energy function $f(x)$, which is assumed to be a function of the z-projection operator only. In this paper, we consider a polynomial form of $f(x)$, although the general case can be treated in a similar fashion. Polynomial terms of the form $f(x) = x^p$ have a natural microscopic form of p-spin interaction of unit strength $H = (2/N)^p \sum \hat{\sigma}_{i_1}^z \hat{\sigma}_{i_2}^z \ldots \hat{\sigma}_{i_p}^z$. For example, a uniform version of the Sherrington-Kirkpatrick model $\sum_{ij} J\sigma_i^z\sigma_j^z = J(\sum_i \sigma_i^z)^2$ corresponds to $f(x) = x^2$. In the same way, a fully connected graph with three-spin interactions $\sum_{ijk} J\sigma_i^z\sigma_j^z\sigma_k^z = J(\sum_i \sigma_i^z)^3$ corresponds to the cubic form of the potential energy $f(x) = x^3$. Quantum dynamics of the fully connected graphs is not a purely theoretical pursuit, it can be implemented experimentally on existing analog quantum annealing hardware [27], reducing the model to local interactions only [28,29]. Moreover, three-spin interactions may be implemented directly in superconducting circuits [30] motivated by the novel physics they could introduce [31].

Without loss of generality, we choose both of the terms in the Hamiltonian Eq. (1) to scale linearly with N. The parameter s in Eq. (1) controls the relative strength of the potential energy and the transverse field and changes from $s = 0$ to $s = 1$ in the course of the quantum annealing algorithm.

For a general function $f(x)$, the Hamiltonian Eq. (1) commutes with $\hat{S}^2 \equiv (\hat{S}^x)^2 + (\hat{S}^y)^2 + (\hat{S}^z)^2$, which is therefore a conserved quantity. In the basis of states, $|S, M\rangle$: $\hat{S}^2|S, M\rangle = S(S+1)|S, M\rangle$, with definite total spin S and its projection on the z axis $M = \{-S, \ldots, S\}$, the matrix elements of the Hamiltonian Eq. (1) are given by the standard spin-S rules,

$$\hat{S}^z|S, M\rangle = M|S, M\rangle, \tag{2}$$

$$\hat{S}^\pm|S, M\rangle = \sqrt{S(S+1) - M(M \pm 1)}|S, M \pm 1\rangle, \tag{3}$$

where we introduced raising and lowering operators, $\hat{S}^\pm = \frac{1}{2}(\hat{S}^x \pm i\hat{S}^y)$. We introduce an integer parameter $K = 0, 1, \ldots, \lfloor N/2 \rfloor$ to label the total spin eigenstates $S = (N/2) - K$. The Hamiltonian Eq. (1) is symmetric with respect to exchanges of pairs of spins $\hat{\sigma}_i \leftrightarrow \hat{\sigma}_j$ and, in fact, with respect to all permutations of spins since such operations do not change the sum over all spins $\hat{S}^{\alpha}$. This symmetry introduces high degeneracy of eigenstates dependent on the total spin S. The subspace with the maximal total spin $S = (N/2)$ or $K = 0$ contains $2S + 1$ nondegenerate states (there are no nontrivial permutations) corresponding to all possible projections of the total spin on the z axis. The states with $K \neq 0$ are highly degenerate, with the degeneracy being determined by the representations of the group of permutations. The eigenstate with a total spin labeled by K has the degeneracy $\binom{N}{K} - \binom{N}{K-1} \sim \exp(NQ_k)$, where $k \equiv (K/N) = \{0, (1/N), (2/N), \ldots, (1/N)\lfloor N/2 \rfloor\}$, which corresponds to the entropy term,

$$Q_k \approx -k \ln k - (1-k)\ln(1-k) + O\left(\frac{\ln N}{N}\right),$$

that has to be added to the free energy of a state with a given energy E and total spin parameter k, $\mathcal{F} = \beta E - Q_k$. Each value of k labels a Hilbert subspace completely disconnected from that labeled by different values of k. This is a manifestation of the spin permutation symmetry that will be violated by the coupling to a thermal bath, which introduces matrix elements between states with different k and effective relaxation in the system.

In this paper, we are interested in a cubic potential energy,

$$f(q) = -c(q - q_{\min})^2\left(q - \frac{3q_{\max} - q_{\min}}{2}\right), \qquad (4)$$

where $q \equiv (2/N)M$, $q = \{-(1-2k), \ldots, (1-2k)\}$. Equation (4) is the most general cubic function with the metastable minimum at $q_{\min}$ and the potential barrier top at $q_{\max}$, where $f(q_{\min}) > f(1)$ ensures that $q = 1$ is the global minimum. Without loss of generality, we can put $c = 1$; the only effect of $c \neq 1$ is to rescale the parameter s in Eq. (1). We are interested in analyzing a computational task, which can be formulated for the model Eq. (1) by defining an appropriate “oracle” [32].

The cubic potential is chosen such that the model demonstrates a first-order phase transition that is unavoidable in the course of quantum annealing [34]. Below, we focus on the open-system quantum annealing in the presence of dissipation and nonzero temperature, which is the case more suitable for implementation on current analog quantum annealers.

A. Quantum annealing computation time

In the course of the QA algorithm, the transverse-field parameter s is varied from $s = 0$ to $s = 1$ with a fixed rate v at a fixed inverse temperature β. The goal is to find the ground state with a probability approaching 1, allowing for repeated runs of the algorithm. Here, we assume that finding any state within the ground-state potential well is sufficient to find the solution (as a local search could allow one to identify the lowest energy state within the well). Given the probability $\mathcal{P}_{\mathrm{GS}}$ of finding the ground state after a single run of duration v^{-1}, the number of runs needed to achieve this goal is $\mathcal{P}_{\mathrm{GS}}^{-1}$. The total computation time is therefore given by

$$\tau \sim v^{-1} \times \mathcal{P}_{\mathrm{GS}}^{-1}. \qquad (5)$$

The goal of this paper is to analyze the scaling of the optimal τ with the number of qubits in the system N characterized specifically by the quantity

$$\xi \equiv \frac{1}{N}\log \tau. \qquad (6)$$

Below, we show that the scaling of ξ in our model is dominated by the exponentially slow quantum tunneling through (or classical escape over) a wide barrier separating the metastable and the ground states. The tunneling rate can be analyzed using the WKB wave functions we introduce in the following sections.

B. WKB wave functions

The dynamics of the model Eq. (1) can be described using a systematic quasiclassical WKB expansion. For an excellent review, see Ref. [18]. In a spin model, this expansion is performed in terms of the small parameter $\varepsilon \equiv 2/N \ll 1$, which is an analog of $\hbar$ in the textbook WKB approach [35]. We consider a wave function in the form

$$\Psi = e^{i(1/\varepsilon)\Phi(q)}$$

and expand it, $\Phi(q) \approx \Phi_0 + \varepsilon\Phi_1 + \varepsilon^2\Phi_2 + O(\varepsilon^2)$. We can further expand the coefficients Φ_i for a small shift of the argument from q to $q \pm \varepsilon$,

$$\Psi_{M\pm1} = \Psi_M e^{\pm i\dot{\Phi}_0}\left(1 + i\frac{\varepsilon}{2}\ddot{\Phi}_0 \pm i\varepsilon\dot{\Phi}_1\right),$$

where $\dot{O} \equiv (dO/dq)$. Substituting this expansion into the Schroedinger equation, we obtain

$$\mathcal{H}\Psi(q) \approx e(q)\Psi(q),$$

where in the main order in ε, the Hamiltonian is diagonal and reads

$$e(q) \approx sf(q) - \frac{1}{2}(1-s)\sqrt{(1-2k)^2 - q^2}\cos\dot{\Phi}_0. \qquad (7)$$

Note that the function $p \equiv \dot{\Phi}_0$ is precisely the canonical momentum conjugate to the coordinate q, $p \to -i(d/dq)$. In other words, the second term above is the quantum kinetic energy, which for $p \ll 1$ corresponds to a particle with a position-dependent mass $m^{-1} \equiv \frac{1}{2}(1-s)\sqrt{(1-2k)^2 - q^2}$ moving in an effective classical potential,

$$U(k,q) = sf(q) - \frac{1}{2}(1-s)\sqrt{(1-2k)^2 - q^2}. \quad (8)$$

Note that the mass of the effective quantum particle increases with increasing q and diverges as $q \to 1$, which affects the efficiency of quantum tunneling into the states near $q = 1$. The effective potential Eq. (8) for different values of s is shown in Fig. 2. In the course of the QA algorithm $s = 0 \to s = 1$, the effective potential deforms from a square-root parabola corresponding to the ground state $(q,k) = (0,0)$ with the maximal spin polarization along the transverse-field direction at $s \to 0$ (see left panel in Fig. 2) and the classical potential corresponding to the ground state $(q,k) = (1,0)$ fully polarized along the axis of quantization $s \to 1$. Note that the initial and final states of the algorithm are characterized by an exponentially small overlap (see Appendix B).

States with small kinetic energy are confined to one of the potential wells, centered around the two minima $q_{\min}^{(L)}(s) < q_{\min}^{(R)}(s)$ of the effective potential $U(k,q)$, which in the course of the evolution as $s \to 1$, approach the metastable $q_{\min}^{(L)} \to q_{\min}$ and ground state $q_{\min}^{(R)} \to 1$ of the classical model, respectively. Confinement of the states is determined by the condition of vanishing classical velocity, $v(k,q,E) \equiv (\partial\mathcal{H}/\partial p) = 0$, which gives

$$-1 \le 2\frac{sf(q) - E}{(1-s)\sqrt{(1-2k)^2 - q^2}} \le 1. \quad (9)$$

The solution of this equation gives the location of the turning points $q_{\mathrm{TP}}^{(1)}(k,E)$ and $q_{\mathrm{TP}}^{(2)}(k,E)$, which limit the classically allowed region. Inverting the secular equation, Eq. (7), for a given energy E, we write the wave function in the leading order in ε,

$$\Psi \sim \exp\left(\frac{i}{\varepsilon}\int_{q_{\mathrm{TP}}^{(1)}}^{q} dq p\right), \quad (10)$$

with

$$p = \arccos 2\frac{sf(q) - E}{(1-s)\sqrt{(1-2k)^2 - q^2}}, \quad (11)$$

when the condition Eq. (9) is satisfied, and

$$p = i\mathrm{arccosh}2\frac{sf(q) - E}{(1-s)\sqrt{(1-2k)^2 - q^2}} \quad (12)$$

otherwise. The latter expression corresponds to the exponentially decaying tail of the wave function extending beyond the classically allowed region into the potential barrier.

C. Quantum phase transition

In the course of the evolution, s: $0 \to 1$, there is a point of zero-temperature discontinuous quantum phase transition, $s = s_{\mathrm{QPT}}$, at which the minimal energies in the two wells, left E_L at $q \sim q_{\min}$ and right E_R at $q \sim 1$, are equal to each other. In the course of the QA algorithm at finite temperature, this transition occurs at a weaker transverse-field strength $s = s_{\mathrm{PT}}(\beta) \ge s_{\mathrm{QPT}}$. The phase transition point can be found from the condition of equal occupation of

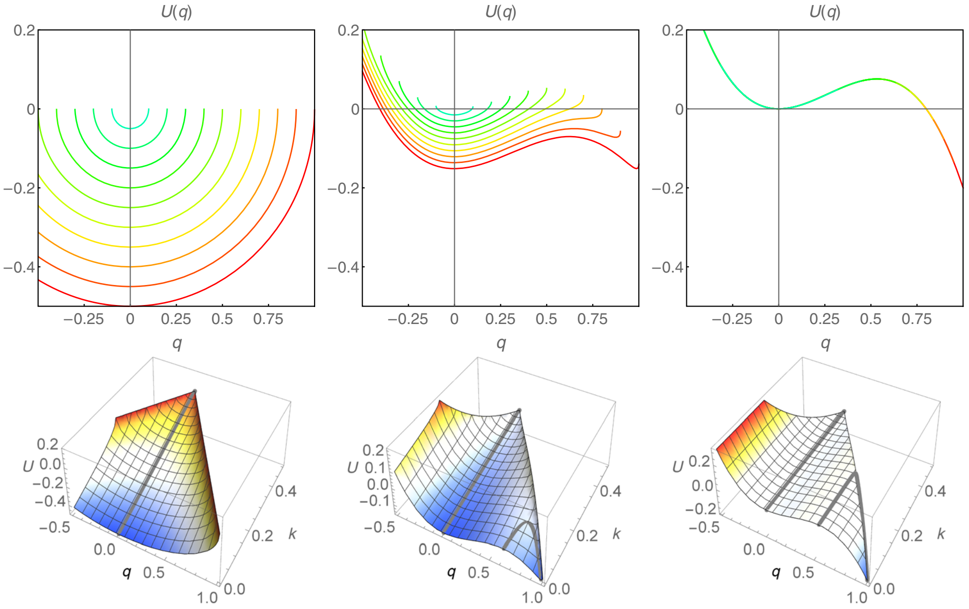

FIG. 2. The effective classical potential is shown for $s = 0$, the pure transverse-field case (left); $s = s_{\mathrm{QPT}} \approx 0.698$, the zero-temperature quantum phase transition point (center); and $s = 1$, the pure classical potential (right). Different lines (bottom to top on all plots) correspond to different values of $k = 0, \ldots, 0.5$ with equal intervals. Note that in the absence of the transverse field, the states with different values of k are degenerate. Here, $q_{\min} = 0$ and $q_{\max} \approx 0.467$.

the two potential wells $\mathcal{P}_L = \mathcal{P}_R$, including the entropy of the states. In the large N limit and at low temperatures $\beta \sim O(N^0)$, we can approximate the occupation number $\mathcal{P}_L = \sum_{E,k} \exp(-N\mathcal{F} - \log Z) \approx \exp(-N\mathcal{F}_L - \log Z)$ by a single dominant term corresponding to a minimum of the free energy. We write the local minimum condition in a potential well as $(\partial\mathcal{F}/\partial k) = 0$ to obtain

$$-\beta\frac{\partial E}{\partial k} = -\frac{\partial Q_k}{\partial k} \approx \ln\frac{k}{1-k}. \tag{13}$$

This determines the k_{min} corresponding to the minimum of the free energy, whereas the optimal energy (principal quantum number) with a given k corresponds to the minimum of the potential $U(k_{\text{min}}, q)$. We neglect the quantization of levels due to finite N for the purpose of this calculation. At $s > s_{\text{PT}}(\beta)$, the metastable (left) potential well is separated from the ground state by a potential barrier with the shape determined by the parameters q_{min} and q_{max}, and the overall scaling $\sim N$. At low temperatures, the relaxation in this model is therefore determined by the rate of transitions between the two wells, which we calculate in the following.

D. Quantum tunneling

In a closed quantum system in the absence of the thermal bath, quantum-mechanical states are coherent superposition of states in the two wells, the so-called Schroedinger cat states. Formally, these states correspond to a coherent (infinite) sum of amplitudes of multiple tunneling events between the wells. In a large system, however, the level splitting Δ corresponding to such a superposition state is exponentially small (in the system size N), and therefore, such coherent dynamics is quickly suppressed by small perturbations, such as the hardware noise. This results in overdamped dynamics characterized by fast intrawell relaxation towards thermal occupation [36] reflected in the level width $W \gg \Delta$ and exponentially rare incoherent tunneling events with the rate $\sim\Delta^2$. This is the regime we consider in this paper. At the same time, we neglect the effect of noise on the tunneling event itself since tunneling is a fast process occurring on the time scale $1/\Omega$, where Ω is the frequency determined by the curvature of the potential. We assume, therefore, that $\Omega \gg W \gg \Delta$. We are interested only in the exponential scaling of the transition rates in this paper, ignoring the renormalization of the preexponential factor that may be substantial in the regime of strong coupling to the environment. Note that the overdamped dynamics and thermalization are expected even in the absence of the coupling to a thermal bath; it can be introduced by a weak disorder in the spin-spin interactions $\delta H = \sum \epsilon_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z$ or even a weak random transverse field [37].

At $s < s_{\text{PT}}(\beta)$, the ground state corresponds to $q = q_{\text{min}}^{(L)}(s)$. As the system goes past the phase transition with growing $s > s_{\text{PT}}(\beta)$, this state becomes metastable. The average transition rate $w \ll W$ across the barrier in the presence of the fast intrawell relaxation W can be obtained by calculating the total current escaping the metastable well [38,39],

$$w \propto \frac{1}{Z}\sum_{k,E} w(k,E)e^{-N(\beta E - Q_k)}. \tag{14}$$

Equation (14) is a thermal average, weighted with the usual Boltzmann factor $e^{-N(\beta E - Q_k)}$, including the entropy Q_k, of $w(k,E) \sim |\Delta(k,E)|^2 \sim e^{-NS_{\text{WKB}}(k,E)}$, the incoherent tunneling amplitude through the barrier of a state with a given energy E and total spin parameter k. The so-called reduced action $S_{\text{WKB}}(k,E)$ can be obtained by matching the quasiclassical wave functions across the barrier region or following the analytical continuation procedure [35],

$$\begin{aligned} S_{\text{WKB}}(k,E) &= -\int_{q_L}^{q_R} dq\, p(q) \\ &= -\int_{q_L}^{q_R} dq \operatorname{arccosh}\frac{2(sf(q) - E)}{(1-s)\sqrt{(1-2k)^2 - q^2}}. \end{aligned}$$

The sum in Eq. (14) can be approximated by its largest term, the rest being exponentially smaller,

$$w \sim \frac{1}{Z}\sum_{k,E} w(k,E)e^{-N(\beta E - Q_k)} \approx e^{-NS}, \tag{15}$$

where the largest term is found by minimizing the action,

$$S = S_{\text{WKB}} + \beta E - Q_k + \frac{1}{N}\ln Z \tag{16}$$

with respect to k and E,

$$T(E) \equiv -\frac{\partial S_{\text{WKB}}}{\partial E} = \beta, \tag{17}$$

$$-\frac{\partial S_{\text{WKB}}}{\partial k} - \beta\frac{\partial E}{\partial k} = -\frac{\partial Q_k}{\partial k} \approx \ln\frac{k}{1-k}. \tag{18}$$

In Eq. (17), fixed k is assumed. Since Q_k is independent of the energy level E, the conditions on the optimal tunneling parameters separate into the standard condition Eq. (17) [requiring the period of motion in the inverted potential $T(E)$ to match the inverse temperature β [39]] and the condition Eq. (18) due to the entropy of states dependent on k, which introduces novel physics in the dynamics of this model. Equations (16)–(18) need to be supplemented with conditions ensuring that the energy and total spin k are conserved in the tunneling event (we emphasize that we neglect here the effect of the thermal bath during the tunneling event). The tunneling from the metastable well has to be at energy and spin values, E and k, at which a state

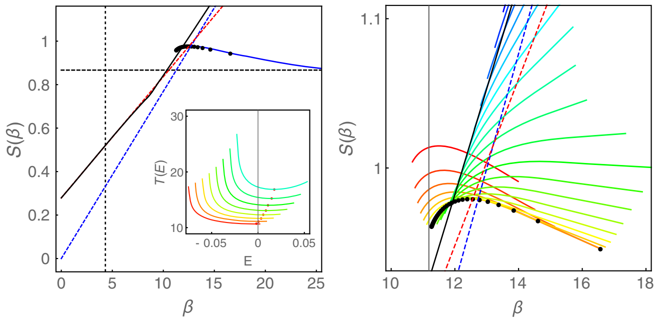

FIG. 3. Tunneling action as a function of inverse temperature β, for $q_{\min} = 0$, $q_{\max} = 0.533$ at $s = 0.85 > s_{\mathrm{QPT}}$. Left: Black dots and the solid blue line fit correspond to an optimal finite-temperature quantum-tunneling action. The horizontal dotted line is the $\beta \to \infty$ limit of the incoherent tunneling action corresponding to the quantum-mechanical tunneling from the lowest level of the metastable well corresponding to $k = 0$. The solid black line corresponds to the optimal action of the classical Glauber dynamics. Blue and red dashed lines show the classical over-the-barrier escape action at $k = 0$ and along the line $q = 1 - 2k$ with the potential maximum at $k_* \approx 0.152$, respectively. The latter, k_*, is the point of inflection, i.e., the point where the minimum of the right potential well merges with the maximum at the barrier top (see Fig. 2). The vertical dashed line corresponds to the temperature-driven phase transition point $\beta_{PT} \approx 4.32$ corresponding to equal occupation of the two potential wells. Inset: $T(E) = -\partial S/\partial E$ at fixed k, which corresponds to the period of quantum-mechanical tunneling trajectory in the imaginary time representation. The dominant contribution comes from tunneling at the energy determined from $T(E) = \beta$ for $\beta > T_{\min}$; at $\beta < T_{\min}$, the transition rate is dominated by the over-the-barrier escape. Different lines from bottom to top correspond to different values of s at fixed $k = 0$. Red dots correspond to $T_{\min}$. Note that the minimum $E_{\min}$: $T(E_{\min}) = T_{\min}$ does not correspond to the highest tunneling energy. This means that the transition from the quantum-tunneling regime to over-the-barrier escape has discontinuous first-order character. Right: Quantum-classical transition region on a larger scale. Black dots correspond to the dominant tunneling action at each β. Different lines show the quantum-mechanical tunneling action for fixed k; colors correspond to growing $0 \leq k \leq 0.152$ (red to blue). The quantum-tunneling process conserves energy and the total spin value k. At $k > k_{\mathrm{LR}}$, the point where $E_L(k_{\mathrm{LR}}) = E_R(k_{\mathrm{LR}})$, the quantum mechanical tunneling process requires the state in the metastable well to be at an energy $E \geq E_R(k)$, which comes with an additional thermal excitation cost $\beta[E_R(k) - E_L(k)]$ in the tunneling action. Therefore, for growing k, the tunneling action $S(\beta)$ resembles linear classical dependence. Solid black and dashed red and blue lines are the same as in the left figure.

exists in the ground state well, i.e., $E \geq \min\{E_R(k)\}$, which is not always satisfied in the system with large entropy of states; i.e., $\mathcal{F}_L < \mathcal{F}_R$ does not necessarily imply $E_L < E_R$.

Equation (17) has a solution in a range of energies E such that $T_{\min} \leq T(E) < \infty$ (see inset in Fig. 3, left). In the case of $\beta > T_{\min}$, the quantum-tunneling process dominates in the sum in Eq. (15). For $\beta < T_{\min}$, there are no solutions to Eq. (17), and therefore, the optimal energy is at the edge of the interval $E = U(q_{\max}) - U(q_{\min})$ corresponding to the height of the barrier. In other words, in this regime, the over-the-barrier escape process dominates, with $\beta \sim T_{\min}$ being the point of a quantum-to-classical phase transition. Note that the global minimum $E_{\min}$ of the function $T(E)$ in the inset in Fig. 3 does not always correspond to the top of the barrier, which means that the quantum-to-classical transition (in the limit $N \to \infty$) has a discontinuous first-order character [19]. Considering Eq. (18), we look for a solution in the interval $0 \leq k \leq k_*$, where k_* is the inflection point of the potential $U(q, k)$ in which the right (the ground-state) potential well disappears (see Fig. 2); since quantum tunneling conserves the total spin k, for it to occur there must exist states with matching k in the ground-state potential well.

The result of this optimization procedure is the optimal action $S(\beta)$ at a fixed s shown in Fig. 3. In the vanishing temperature limit $\beta \to \infty$, where the effect of entropy on the occupation of levels is negligible, $S(\beta)$ corresponds to the quantum tunneling from the lowest energy level in the metastable well corresponding to $k = 0$ (horizontal dashed line in Fig. 3, left). As temperature increases (β decreases), the entropy starts playing a role in dynamics, and $S(\beta)$ increases up to some maximum value (see blue solid line in Fig. 3, left), in contrast to the usual (nondegenerate) case [39] where the quantum-tunneling rate increases monotonously with increasing temperature β^{-1}. This is a result of the entropy providing high statistical weight to the state with a suboptimal tunneling rate. This behavior is natural because the transverse field splits the degeneracy of the metastable state at $q = q_{\min}$ favoring the state with maximal total spin, corresponding to $k = 0$. At the same time, the transverse field provides maximal quantum kinetic energy (minimal mass in the quadratic approximation) to the same state, as can be seen from Eq. (7), whereas the finite temperature due to the effect of entropy favors the states with $1/2 \geq k > 0$ (at $\beta \to 0$), which have lower kinetic energy (higher mass) and correspond to the suboptimal tunneling rate through the classical potential $sf(q)$. The result is $S(\beta)$ increasing with decreasing β up to the regime of the transition into the classical escape at high temperatures $\beta \ll T_{\min}$. The classical over-the-barrier excitation is described by the thermal excitation rate $w_{\mathrm{cl}} \sim Z^{-1} \exp[-N(\beta E - Q_k)]$. Note that the classical process is driven by the Glauber dynamics of the spins due to the effect of the thermal bath. This process does not conserve the total spin k, and therefore, the optimal classical trajectory is determined by the saddle point of the free energy (including the entropy) in the 2D space of (k, q). Entropy provides an additional cost, reducing the transition rate, which is reflected in the finite offset of the dependence of the classical transition rate on β as $\beta \to 0$ (see black solid line in Fig. 3, left; see Appendix A for more details). Blue and red dashed lines in Fig. 3 indicate the linear in β dependence of the energy cost of the over-the-barrier excitation at $k = 0$ and along the line of $q = 1 - 2k$ with the potential maximum at $k \approx k_*$ [the latter is the inflection point of the potential $U(k, q)$], respectively, which correspond to the low-temperature potential-energy-dominated regime and the high-temperature entropy-dominated regime [40].

The steepest descent approximation Eq. (15) is applicable as long as the preexponential factors in the sum are nondivergent, which is true away from $\beta \sim T_{\min}$, a phase transition point from the quantum-tunneling regime to the classical over-the-barrier escape regime [39]. On general grounds, we expect the action to be continuous even in the case where this quantum-to-classical transition is of the first order, with the discontinuity occurring in the derivative of the action [19]. Therefore, we expect Fig. 3 to provide a qualitatively correct dependence of $S(\beta)$ in the whole range of inverse temperatures β.

III. QUANTUM ANNEALING COMPUTATION TIME

We now turn to the calculation of the computation time defined in Eq. (6). The quantum mechanical tunneling rate vanishes as $s \to 1$ (at very low temperatures of interest, here the over-the-barrier transition rate is also weak). Therefore, there exists a point $s = s_F$ in the course of the sweep of the transverse field where the relaxation time $\sim w^{-1}(s_F)$ required to achieve thermal distribution becomes longer than the length of the algorithm, $w^{-1}(s_F) \sim v^{-1}$. In other words, the computation will be finished before the thermal equilibrium is reached. The system effectively freezes the values of the occupation numbers of the left $\mathcal{P}_L$ and right $\mathcal{P}_R$ wells at the last point (in the course of the algorithm) where the interwell transition process was still fast enough (the intrawell relaxation may still be efficient at $s > s_F$). We call this the freezing point. Sudden freezing of quantum dynamics in the course of quantum annealing is typical for the algorithms relying on quantum tunneling between states separated by large Hamming distance [3,41–44]. The computation time of the algorithm (its exponential scaling) is therefore determined by the thermal (equilibrium) occupation probability $\mathcal{P}_{\mathrm{GS}} = \mathcal{P}_R(s_F)$, and the equilibration time $w^{-1}(s_F)$ at the freezing point,

$$\xi \equiv \frac{1}{N} \log \tau \approx \frac{1}{N} \log \left[e^{NS} (1 + e^{N(\mathcal{F}_R - \mathcal{F}_L)}) \right], \qquad (19)$$

where S is the optimal quantum-tunneling action given by Eq. (16), with $s = s_F$. The optimal computation time of the quantum annealing is found by minimizing with respect to the location of s_F and the temperature β at which the computation is performed. The point of the phase transition, $\beta = \beta_{\mathrm{PT}}$ and $s_F = s_{\mathrm{PT}}$, respectively, defined by $\mathcal{F}_R - \mathcal{F}_L = 0$, separates two scaling regimes of the computation time,

$$\xi \approx \begin{cases} S + \mathcal{F}_R - \mathcal{F}_L & \beta < \beta_{\mathrm{PT}}(s) \\ S & \beta > \beta_{\mathrm{PT}}(s). \end{cases} \qquad (20)$$

The high-temperature limit ($\beta \to 0$) of this expression is given by the entropy difference between the two wells. This is the limit of a local exhaustive search, which is a bound on the computation time of the algorithm, and we need to compare it to the optimal computation time $\xi(\beta)$ to make sure we find the global optimum. We first consider $\beta < \beta_{\mathrm{PT}}$, assuming the free energy of the system demonstrates two minima, $\mathcal{F}_R$ and $\mathcal{F}_L$. In the quantum-tunneling regime, both $S(\beta)$ and $\mathcal{F}_R - \mathcal{F}_L$ decrease with growing β, and therefore, $\xi(\beta)$ is monotonously decreasing as well. In the classical regime, $S + \mathcal{F}_R - \mathcal{F}_L$ is also a monotonous function, which depending on competition between $S(\beta)$ and $\mathcal{F}_R - \mathcal{F}_L$, can be either increasing, in which case $\beta \to 0$ is the optimal classical computation regime (i.e., the local exhaustive search limit), or decreasing towards the critical point $\beta = \beta_{\mathrm{PT}}$. Therefore, we need to analyze the performance in the regime $\beta > \beta_{\mathrm{PT}}$ and $s > s_{\mathrm{PT}}$ and compare it to a local exhaustive search. The optimal computation time in this regime [see Eq. (20)] is determined by the minimum of the transition action S with respect to β and s.

Because of the concave dependence of $S(\beta)$ on the inverse temperature, as shown in Fig. 3 (left), the minimum of the action with respect to β corresponds to one of the edges of the inverse temperature interval, i.e., either $\beta = \beta_{\mathrm{PT}}$ or $\beta \to \infty$. The global minimum is therefore the smallest of $S[\beta_{\mathrm{PT}}(s_F), s_F]$ and $S(\infty, s_F)$. The minimum with respect to s_F (and therefore the global minimum) is determined by comparing these two functions.

A typical critical line is shown in the left panel of Fig. 4. The inverse critical temperature $\beta_{\mathrm{PT}}(s)$, blue solid line in Fig. 3 (left), diverges at the point of the quantum phase transition in the course of the algorithm, $s = s_{\mathrm{QPT}}$ (vertical dashed line in Fig. 4), and at $s > s_{\mathrm{QPT}}$ it monotonously decreases, with s approaching the classical transition temperature at $s = 1$ (horizontal dashed line in Fig. 4). The right panel of Fig. 4 shows the over-the-barrier escape action at the given critical temperature $\beta_{\mathrm{PT}}(s)$, blue (upper) solid line. The blue dots in the left and right panels correspond to the same values of s. The classical action, following the behavior of the inverse critical temperature, diverges at the point of the quantum phase transition $s = s_{\mathrm{QPT}}$ and decreases with s in the course of the algorithm approaching the value corresponding to the classical model $s = 1$, shown as the blue (upper) dashed line. The latter is the optimum (due to the highest critical temperature) and therefore defines the optimal computation time of the classical simulated annealing algorithm (see Appendix A for more details). The optimum of the quantum action, at vanishing temperature $\beta \to \infty$, is given by the tunneling from the bottom of the metastable well. In this limit, the entropy does not affect the occupation of the energy levels in the course of the algorithm. We assume that the

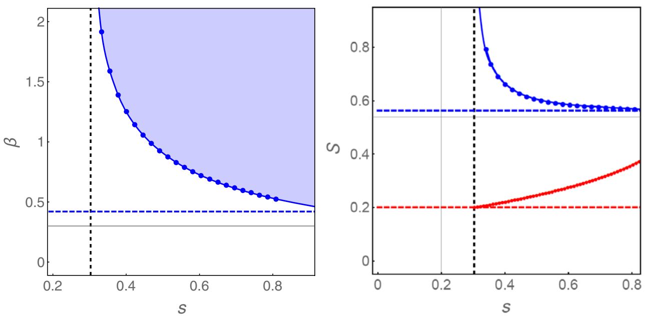

FIG. 4. Left panel: Inverse critical temperature dependence on the transverse-field strength $\beta_{\mathrm{PT}}(s)$ for $q_{\min} = 0.88$, $q_{\max} = 0.955$. The horizontal dashed line corresponds to the classical transition temperature, and the vertical dashed line marks the point of the quantum phase transition. Right panel: Comparison of the vanishing temperature quantum-tunneling action to that of the classical over-the barrier escape. The blue line shows optimal over-the-barrier escape action along the critical lineshown in the left panel. Blue points correspond to the same values of s as in the left figure. The horizontal blue line corresponds to the transition rate at the critical temperature in the classical model, $s = 1$. The vertical line corresponds to the point of zero-temperature quantum phase transition. The red line corresponds to the vanishing temperature limit of the quantum-tunneling action. The red horizontal dashed line corresponds to the minimum of the quantum-tunneling action at the quantum phase transition point. The local exhaustive search corresponds to $\xi = (1/N)\log\tau_{\mathrm{es}} = Q_{\mathrm{cl}}(q_{\min}) \approx 0.227$, larger than the quantum annealing time.

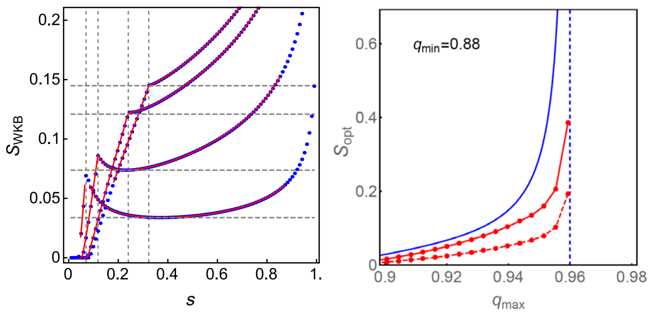

FIG. 5. Left panel: Quantum mechanical tunneling action at vanishing temperature in the incoherent regime as a function of the transverse-field parameter s. Different curves correspond to different values of $q_{\max} = 0.929$, 0.946, 0.958, 0.961 (order is from bottom to top on the right side of the plot) with fixed $q_{\min} = 0.9$. Note that the optimal tunneling rate emphasized by the horizontal dashed lines does not always correspond to the phase-transition point denoted by vertical dashed lines. Right panel: Optimal action for SA (blue) and QA (red) at $q_{\min} = 0.88$ for different values of $q_{\max}$ [see Eqs. (19) and (20) and the text for details]. The dashed red line corresponds to the coherent quantum-tunneling action (scaling as 1/2 of the action in the incoherent regime). The vertical dashed line corresponds to $q_{\min} = \frac{3}{2} q_{\max}$, at which point the classical potential has a degenerate ground state at $q = q_{\min}$ and $q = 1$. At this point, both the SA action and the QA action diverge; however, the QA action diverges logarithmically slow, in contrast to the SA action. For the parameters chosen, the local exhaustive search corresponds to $\xi = (1/N)\log\tau_{\mathrm{es}} = Q_{\mathrm{cl}}(q_{\min}) \approx 0.227$, which is above the value of the QA action in the figure, except for the close vicinity of the divergence point, i.e., QA is faster than SA and an exhaustive search for most of the parameter values.

temperature is still high enough such that there is sufficiently fast intrawell relaxation. Figure 4 (right panel) shows $S(\infty, s)$, red (lower) solid line, which assumes the minimal value at the quantum phase transition point $s = s_{\mathrm{QPT}}$. Note that this is not true for all the parameter values; instead, the minimum of $S(\infty, s)$ often occurs at some $s > s_{\mathrm{QPT}}$, and its value at this point can be as much as 2 times smaller than the value at $s = s_{\mathrm{QPT}}$ (see Fig. 5, left panel).

Figure 4 (right panel) demonstrates that the quantum-tunneling process may be more efficient than over-the-barrier escape. Both classical and quantum transition rates, and therefore the corresponding computation times, scale exponentially with the system size, $\tau \propto \exp(\alpha N)$, yet the coefficient in the exponent is smaller in the case of QA as compared to the classical simulated annealing, which corresponds to a polynomial speedup. Note that it is important to compare the computation times in Fig. 4 with that of the local exhaustive search in the interval $(q_{\min}, 1)$, which is the high-temperature limit of SA. The corresponding computation time is given by the entropy $\xi = Q_{\mathrm{cl}}(q_{\min})$, which for parameters considered in Fig. 4, $Q_{\mathrm{cl}}(0.88) \approx 0.227$, scales worse than the QA computation time shown by red dots in Fig. 4 (right panel). We further compare the optimal performance of the open-system quantum annealing S_{opt} to simulated annealing for a wide range of barrier shapes within the cubic model Eq. (4) by varying the location of the metastable minimum $q_{\min}$ and the barrier top $q_{\max}$ in Eq. (4) (see right panel in Figs. 5 and 6). QA outperforms SA in a range of the parameter space where the potential barrier separating the metastable and the ground states is small. Note that the origin of the speedup in the case considered here is distinct from the standard intuition of thin and tall barriers favoring quantum tunneling (see Ref. [13] for refining the standard intuition) since the shape of the potential is cubic throughout the range of parameters shown in Fig. 6. Instead, the quantum algorithm turns out to be more efficient because it proceeds along a path with lower entropic cost than the path that SA takes. This is a result of the transverse field lifting the degeneracy of the metastable state in the quantum case. Therefore, the smallness of the barrier required for the speedup in this case is determined by the comparison of the quantum-tunneling action and the SA action, including entropy cost of over-the-barrier escape, the latter being a combination of $\beta_{\mathrm{PT}} U$ and the additional entropic cost $Q_{\mathrm{cl}}(q_{\min}) - Q_{\mathrm{cl}}(q_{\mathrm{top}})$, where q_{top} is the maximum of the free energy (see Appendix A for details). Note also that the numerical value of the ratio of the logarithms of the normalized

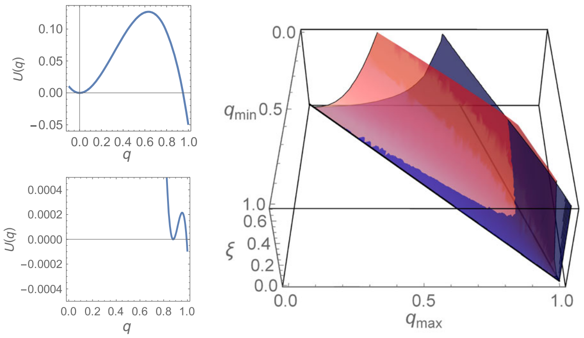

FIG. 6. Left panels: Potential barrier between the metastable and the ground state, Eq. (4), corresponding to $s = 1$ for two pairs of values $(q_{\min}, q_{\max})$ equal to (0, 0.633) (upper plot) and (0.88, 0.955) (lower plot). Right panel: Scaling exponent of the optimal computation time of quantum annealing (red) [see Eqs. (6) and (19)] and simulated annealing (blue) given by $(1/N)\log\tau$, with τ given by Eq. (A2) as a function of the location of the metastable minimum $q_{\min}$ and the top of the barrier $q_{\max}$. The lower of the two surfaces corresponds to the shorter computation time. QA is advantageous for a range of parameters corresponding to sufficiently narrow potential barriers. Solid lines in the plane $S = 0$ correspond to $q_{\min} = q_{\max}$ and $q_{\min} = \frac{3}{2} q_{\max}$ (under both surfaces in the figure) outlining the range of possible potentials in the cubic model Eq. (4).

computation times $\xi \equiv (1/N)\log\tau$ for QA and SA can be substantial [see Fig. 5 (right panel)].

IV. DISCUSSION

In this paper, we consider a model problem of open-system quantum annealing that allows analytical investigation and, at the same time, demonstrates key features of complex optimization problems, including the discontinuous first-order phase transition and the exponential degeneracy of the metastable state. We demonstrate that, for problems with extensive degeneracy, the SA algorithm relies on over-the-barrier escape at very low temperatures, $\beta \sim O(1)$; as a result, the SA computation time is exponential in N. In addition, the classical over-the-barrier escape rate is reduced because of an additional entropic barrier, the difference in the statistical weights of the metastable well and the states near the top of the barrier. At the same time, we show that a computational advantage can be gained using open-system quantum annealing, which exploits the effects of thermally assisted quantum tunneling and relaxation. In the course of QA, the applied transverse field splits the degeneracy of the metastable state and provides the nondegenerate low-energy states with the highest quantum kinetic energy; in other words, the transverse field favors the quantum-tunneling trajectories involving nondegenerate states. In this way, the QA algorithm naturally circumvents the entropic barrier and thus gains a clear advantage over SA. This is a novel feature in the quantum-tunneling process caused by the entropy of the metastable states, manifested particularly in the tunneling rate decreasing with increasing temperature. As a consequence, the optimal quantum annealing regime corresponds to vanishing temperature; i.e., raising the temperature reduces the efficiency of QA [45]. The faster tunneling with lower temperature contrasts the standard intuition that raising temperature should always improve transition rates. In this paper, we provided an important and very widespread counterexample to this standard intuition. We also found that at low temperatures, the optimal quantum-tunneling rate does not always correspond to the point of the phase transition s_{QPT}; in fact, tunneling at $s > s_{\mathrm{QPT}}$ can have a substantially higher rate, which can be exploited in conjunction with the noise-induced thermalization to improve the performance of the algorithm. In the mean-field model considered in this paper, the comparison of the quantum annealing and simulated annealing comes down to the numerical coefficient in the scaling of the computation time with N. We demonstrated that optimal QA could outperform SA in a certain parameter range of our model characterized by small potential barriers. This is in spite of the constrained nature of the quantum-tunneling process due to the conservation laws, as opposed to the unconstrained classical Glauber dynamics of SA.

We also identified key features of the model affecting the performance of the QA, specifically, the quantum fluctuations strength in the vicinity of the phase transition point and the diverging mass in the quantum kinetic energy, which both strongly affect the efficiency of quantum tunneling in the course of the algorithm. In this respect, it would be interesting to use methods developed in this paper to explore QA in mean-field models with different types of driver Hamiltonians whose ground states are not simple product states where ferromagnetic order competes with the transverse (XY) ferromagnetism or superfluidity [46–49].

Note that the model of escape from a highly degenerate potential well considered in this paper is quite generic, as it is natural to expect that the applied transverse field will favor the trajectories involving nondegenerate states, thus circumventing the entropic barrier, which slows down classical stochastic dynamics. This is because the states most strongly affected by the transverse field both acquire the largest energy shifts and the highest kinetic energy and hence the highest transition rates. This suggests a speculation that optimization problems in which entropy is a dominant factor yet, at the same time, which are characterized by a potential energy landscape that can be exploited for a more efficient search (Grover's unstructured search problem [50] with an added potential landscape), may represent a class of problems where quantum annealing could have a computational advantage over simulated annealing.

ACKNOWLEDGMENTS

This work is supported by NASA under Grant No. NNX12AK33A and in part by the Office of the

Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IAA 145483, and by the AFRL Information Directorate under Grant No. F4HBKC4162G001.

APPENDIX A: SIMULATED ANNEALING

We discuss the performance of simulated annealing on the model, Eqs. (1) and (4), in the purely classical case of $s = 1$ (see the rightmost panel in Fig. 2). The SA algorithm is realized by starting with the infinite temperature limit $\beta \to 0$, i.e., equal occupation of all states, and reducing the temperature to zero. For simplicity (and without loss of generality), we consider an algorithm where β changes linearly in time from 0 to a very large value with a fixed rate v.

The time it takes to find the ground state with probability approaching around 1, allowing for repeated runs of the algorithm, is given, as in the quantum case, by

$$\tau \sim v^{-1} \times \mathcal{P}_{\mathrm{GS}}^{-1}.$$

Simulated annealing relies on the system reaching thermal equilibrium throughout at least part of the algorithm, such that the ground-state occupation is given by the Gibbs distribution. In problems with a free energy barrier separating the initial and the ground state, the system's relaxation time is dominated by the over-the-barrier escape probability with $\mathcal{F}(q) = \beta E(q) + Q_{\mathrm{cl}}(q)$,

$$w(\beta) \sim \exp\left[-N(\mathcal{F}(q_{\mathrm{top}}) - \mathcal{F}(q_{\mathrm{min}}))\right],$$

given by the statistical weight of the escape trajectory [26], where we include the entropy of the classical state with magnetization q given by $Q_{\mathrm{cl}}(q) \approx -[(1+q)/2]\ln[(1+q)/2] - [(1-q)/2]\ln[(1-q)/2]$, and q_{top} is the maximum of $\mathcal{F}(q)$. Here, $w(\beta)$ reduces exponentially with decreasing temperature (growing β), and therefore, in analogy with the quantum case considered in the main text, there exists a freezing point $\beta = \beta_F$ in the course of the sweep of the inverse temperature where the relaxation time $\sim w^{-1}(\beta_F)$ required to achieve thermal distribution becomes longer than the length of the algorithm, $w^{-1}(\beta_F) \sim v^{-1}$. After this point, the values of the occupation numbers of the left $\mathcal{P}_L$ and right $\mathcal{P}_R$ potential wells are effectively frozen. Therefore, the computation time is determined by two quantities calculated at the freezing point, the occupation of the ground-state potential well $\mathcal{P}_{\mathrm{GS}} = \mathcal{P}_R(\beta_F)$ and the relaxation time at the freezing point $w^{-1}(\beta_F)$,

$$\tau \approx e^{N[\mathcal{F}(q_{\mathrm{top}}) - \mathcal{F}(q_{\mathrm{min}})]}\left(1 + e^{N[\mathcal{F}(1) - \mathcal{F}(q_{\mathrm{min}})]}\right), \qquad \text{(A1)}$$

where we keep only the main order in the limit $N \to \infty$ such that $\mathcal{P}_{\mathrm{GS}}(\beta_F) \approx \exp[-\mathcal{F}(1)]/[\exp(-\mathcal{F}(1)] + \exp[-\mathcal{F}(q_{\mathrm{min}})]$. The optimal computation time can be obtained by minimizing with respect to the inverse temperature at the freezing point, $(\partial/\partial\beta_F)[(1/N)\log_2\tau] = 0$. This derivative is discontinuous at the point of the phase transition β_{PT}, where $\mathcal{F}_R - \mathcal{F}_L = 0$. The computation time is an increasing function at $\beta_F > \beta_{\mathrm{PT}}$, as it is dominated by the decreasing transition rate, the prefactor in front of the curly brackets in Eq. (A1). However, it can be either a monotonously decreasing or increasing function at $\beta_F < \beta_{\mathrm{PT}}$ depending on the competition between the prefactor and the exponents in the brackets in Eq. (A1). Therefore, the global minimum of $\tau(\beta_F)$ corresponds to the smallest value out of $\tau(\beta_{\mathrm{PT}})$ and $\tau(0) \approx 2^N$, in the decreasing and increasing cases, respectively. The latter corresponds to the exhaustive search, i.e., 2^N repetitions of infinitely fast SA. The ground state at $q = 1$ is unique, $Q(1) \approx 0$; therefore, the point of the classical phase transition is at $\beta_{\mathrm{PT}} = [Q(q_{\mathrm{min}})/E(q_{\mathrm{min}}) - E(1)]$. Therefore, the optimal SA computation time corresponds to the smaller of the two values,

$$\tau \to \left[\begin{array}{c} \exp\left[N\left(\frac{E(q_{\mathrm{top}}) - E(q_{\mathrm{min}})}{E(q_{\mathrm{min}}) - E(1)} Q(q_{\mathrm{min}}) + \delta Q\right)\right], \\ 2^N, \end{array}\right. \qquad \text{(A2)}$$

where $\delta Q \equiv Q(q_{\mathrm{min}}) - Q(q_{\mathrm{top}})$. Note that the entropy associated with the metastable state causes a very low transition temperature $\beta_{\mathrm{PT}} \sim O(1)$ and gives rise to an additional statistical factor $\exp(\delta Q)$ slowing down the transitions over the barrier. This additional factor appears as a prefactor in the Kramers rate calculation [26]; in the model considered here, this factor is exponential and needs to be included to correctly describe the scaling of the classical transition rate with N.

APPENDIX B: EIGENSTATES OVERLAP

One of the characteristics associated with complexity of a given problem for the quantum annealing algorithm is the overlap of the eigenstate wave functions at the beginning and the end of the algorithm. We analyze it for our model, Eqs. (1) and (4). The initial state $s = 0$ is characterized by the maximal x projection of the total spin,

$$\hat{S}^x \left|\frac{N}{2} - K\right\rangle_x = \left(\frac{N}{2} - K\right)\left|\frac{N}{2} - K\right\rangle_x.$$

The overlap of this state with the solution of the classical problem, the state fully polarized along z axis, is

$${}_x\langle N/2 - K|\vec{0}\rangle = \left[\frac{1}{2^N}\binom{N}{K}\right]^{1/2}.$$

[1] A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zwerger, *Dynamics of the Dissipative Two-State System*, Rev. Mod. Phys. **59**, 1 (1987).

[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, 2011).

[3] S. Boixo, V. N. Smelyanskiy, A. Shabani, S. V. Isakov, M. Dykman, V. S. Denchev, M. Amin, A. Smirnov, M. Mohseni, and H. Neven, *Computational Role of Collective Tunneling in a Quantum Annealer*, arXiv:1411.4036.

[4] J. W. Britton, B. C. Sawyer, A. C. Keith, C.-C. Joseph Wang, J. K. Freericks, H. Uys, M. J. Biercuk, and J. J. Bollinger, *Engineered Two-Dimensional Ising Interactions in a Trapped-Ion Quantum Simulator with Hundreds of Spins*, Nature (London) **484**, 489 (2012).

[5] T. Kadowaki and H. Nishimori, *Quantum Annealing in the Transverse Ising Model*, Phys. Rev. E **58**, 5355 (1998).

[6] S. Geman and D. Geman, *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721 (1984).

[7] Z. Jiang, C. Papadimitriou, *Computational Complexity* (Addison-Wesley, Reading, Massachusetts, 1994).

[8] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, *Quantum Annealing: A New Method for Minimizing Multidimensional Functions*, Chem. Phys. Lett. **219**, 343 (1994).

[9] J. Brooke, D. Bitko, F. T. Rosenbaum, and G. Aeppli, *Quantum Annealing of a Disordered Magnet*, Science **284**, 779 (1999).

[10] E. Farhi, J. Goldstone, and S. Gutmann, *Quantum Adiabatic Evolution Algorithms versus Simulated Annealing*, arXiv: quant-ph/0201031.

[11] G. E. Santoro, R. Martonak, E. Tosatti, and R. Car, *Theory of Quantum Annealing of an Ising Spin Glass*, Science **295**, 2427 (2002).

[12] V. Bapst, L. Foini, F. Krzakala, G. Semerjian, and F. Zamponi, *The Quantum Adiabatic Algorithm Applied to Random Optimization Problems: The Quantum Spin Glass Perspective*, Phys. Rep. **523**, 127 (2013).

[13] S. Muthukrishnan, T. Albash, and D. A. Lidar, *Tunneling and Speedup in Quantum Optimization for Permutation-Symmetric Problems*, arXiv:1511.03910.

[14] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven, *What is the Computational Value of Finite Range Tunneling?*, arXiv: 1512.02206 [Phys. Rev. X (to be published)].

[15] Note that the 2D spin-glass problem at finite temperature may not be a good benchmark for the performance of the quantum annealing, as the spin-glass transition in this case occurs only at zero temperature; therefore, the phase space of the problem is expected to be rather simple [16].

[16] H. G. Katzgraber, F. Hamze, and R. S. Andrist, *Glassy Chimeras Could Be Blind to Quantum Speedup: Designing Better Benchmarks for Quantum Annealing Machines*, Phys. Rev. X **4**, 021008 (2014); M. Weigel, H. G. Katzgraber, J. Machta, F. Hamze, and R. S. Andrist (Octomore Collaboration), Phys. Rev. X **5**, 019901(E) (2015).

[17] B. Heim, T. F. Ronnow, S. V. Isakov, and M. Troyer, *Quantum Versus Classical Annealing of Ising Spin Glasses*, Science **348**, 215 (2015).

[18] P. A. Braun, *Discrete Semiclassical Methods in the Theory of Rydberg Atoms in External Fields*, Rev. Mod. Phys. **65**, 115 (1993).

[19] S. A. Owerre and M. B. Paranjape, *Macroscopic Quantum Tunneling and Quantum-Classical Phase Transitions of the Escape Rate in Large Spin Systems*, Phys. Rep. **546**, 1 (2015).

[20] V. Bapst and G. Semerjian, *On Quantum Mean-Field Models and Their Quantum Annealing*, J. Stat. Mech. (2012) P06007.

[21] B. Seoane and H. Nishimori, *Many-Body Transverse Interactions in the Quantum Annealing of the p-spin Ferromagnet*, J. Phys. A **45**, 435301 (2012).

[22] S. Suzuki, H. Nishimori, and M. Suzuki, *Quantum Annealing of the Random-Field Ising Model by Transverse Ferromagnetic Interactions*, Phys. Rev. E **75**, 051112 (2007).

[23] Y. Seki and H. Nishimori, *Quantum Annealing with Antiferromagnetic Fluctuations*, Phys. Rev. E **85**, 051112 (2012).

[24] S. V. Isakov, G. Mazzola, V. N. Smelyanskiy, Z. Jiang, S. Boixo, H. Neven, and M. Troyer, *Understanding Quantum Tunneling through Quantum Monte Carlo Simulations*, arXiv:1510.08057.

[25] Z. Jiang, V. N. Smelyanskiy, S. V. Isakov, S. Boixo, G. Mazzola, M. Troyer, and H. Neven, *Scaling Analysis and Instantons for Thermally-Assisted Tunneling and Quantum Monte Carlo Simulations*, arXiv:1603.01293.

[26] P. Hanggi, P. Talkner, and M. Borkovec, *Reaction-Rate Theory: Fifty Years after Kramers*, Rev. Mod. Phys. **62**, 251 (1990).

[27] D. Venturelli, S. Mandrà, S. Knysh, B. O'Gorman, R. Biswas, and V. Smelyanskiy, *Quantum Optimization of Fully Connected Spin Glasses*, Phys. Rev. X **5**, 031040 (2015).

[28] R. Babbush, B. O'Gorman, and A. Aspuru-Guzik, *Resource Efficient Gadgets for Compiling Adiabatic Quantum Optimization Problems*, Ann. Phys. (Amsterdam) **525**, 877 (2013).

[29] W. Lechner, P. Hauke, and P. Zoller, *A Quantum Annealing Architecture with All-to-All Connectivity from Local Interactions*, Sci. Adv. **1**, e1500838(2015).

[30] Sam Y. Cho and M. D. Kim, *Macroscopic Many-Qubit Interactions in Superconducting Flux Qubits*, Phys. Rev. B **77**, 212506 (2008).

[31] Y.-X. Chen and S.-W. Li, *Quantum Refrigerator Driven by Current Noise*, Europhys. Lett. **97**, 40003 (2012).

[32] From the point of view of computation, the formulation of the problem by Eqs. (1) and (4) is incomplete since the optimum of energy, and therefore the solution, is trivially at $q = 1$, whereas a practical problem is expected to have an unknown solution. Given a problem with such an unknown solution we could, in principle, map it onto Eqs. (1) and (4) by applying a unitary transformation (an analog of an oracle in computation theory), where the transformation matrix is unknown. This transformation maps the solution, an unknown classical bit string (a set of Ising spin-up or spin-down states), onto another classical bit string, the fully polarized state with $q = 1$, which means it can be written in terms of spin-flip operators as follows: $U = \sum_{i=1}^{N} (-1)^{a_i} \sigma_i^x$, where $a_i = \{0, 1\}$ is a set of unknown numbers encoding the bit flips necessary to map the two classical states onto each other. Since the oracle operator commutes with the

transverse-field term in Eq. (1), the quantum annealing dynamics in the practical problem system is identical to the one defined by Eqs. (1) and (4) modulo relabeling of spin directions. Therefore, the complexity of the quantum annealing algorithm computed for the problem defined by Eqs. (1) and (4) is the same as for the original problem with a minimum at an unknown bit string. In this way, the quantum annealing considered here is a general-purpose algorithm that does not rely on the low-dimensional structure of the problem. This is in contrast to, for example, the quantum approximate optimization algorithm suggested in Ref. [33], which relies on the specific structure of the phase space of the problem.

[33] E. Farhi, J. Goldstone, and S. Gutmann, *A Quantum Approximate Optimization Algorithm*, arXiv:1411.4028.

[34] Of specific interest is the spin model with $p = 3$. Models with $p > 3$ can be solved in polynomial time by avoiding the first-order phase transition using advanced driver Hamiltonian, b-spin ($b \geq 2$) interactions inducing transverse (XY) ferromagnetism or antiferromagnetism, in addition to the standard transverse-field term. In this case, arriving at the ground state at the end of the QA evolution does not require extensive quantum tunneling through a barrier. Therefore, a classical model algorithm such as the spin-vector Monte Carlo method, which could follow the deformation of the effective quasiclassical potential imitating the course of QA, is also not expected to encounter a bottleneck requiring exponential computation time. In the case of $p = 3$, the first-order phase transition cannot be avoided in this way; as a result, an exponentially weak tunneling process is critical for finding the ground state in this system. The quantum-tunneling regime in this case cannot be described by a classical model.

[35] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Elsevier, United Kingdom, 1981).

[36] P. J. Crowley and A. Green, *An Anistropic Landau-Lifshitz-Gilbert Model of Dissipation in Qubits*, arXiv:1503.00651.

[37] B. Zhao, M. C. Kerridge, and D. A. Huse, *Three Species of Schroedinger Cat States in an Infinite-Range Spin Model*, Phys. Rev. E **90**, 022104 (2014).

[38] D. A. Garanin, X. Martínez Hidalgo, and E. M. Chudnovsky, *Quantum-Classical Transition of the Escape Rate of a Uniaxial Spin System in an Arbitrarily Directed Field*, Phys. Rev. B **57**, 13639 (1998).

[39] E. M. Chudnovsky and J. Tejada, *Macroscopic Quantum Tunneling of the Magnetic Moment* (Cambridge University Press, Cambridge, England, 2005).

[40] Note that in the case of Grover's unstructured search problem corresponding to a completely flat potential $U(k, q) = 0$ everywhere except $U(0, 1) = 1$, the offset of the classical action in Fig. 3 corresponds to the classical complexity (computation time) of Grover's problem with respect to simulated annealing, 2^N. Whereas in the presence of the potential profile, $U(k, q) \neq 0$, which provides structure to the problem, the complexity may be lower.

[41] V. N. Smelyanskiy, D. Venturelli, A. Perdomo-Ortiz, S. Knysh, and M. I. Dykman, *Quantum Annealing via Environment-Mediated Quantum Diffusion*, arXiv:1511.02581.

[42] T. Albash, S. Boixo, D. A. Lidar, and P. Zanardi, *Quantum Adiabatic Markovian Master Equations*, New J. Phys. **14**, 123016 (2012).

[43] M. H. Amin, *Searching for Quantum Speedup in Quasistatic Quantum Annealers*, Phys. Rev. A **92**, 052323 (2015).

[44] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk *et al.*, *Quantum Annealing with Manufactured Spins*, Nature (London) **473**, 194 (2011).

[45] Note that vanishing temperature is limited from below only by the condition that local relaxation within a potential well due to coupling to a thermal bath is not exponentially slow (in other words, it is faster than tunneling). For the weak coupling regime of an Ohmic bath, for which a single-qubit relaxation rate is proportional to the temperature, an exponentially small rate means exponentially small temperature. At the same time, a significant change of QA computation time requires a change in temperature of order one, determined by the temperature dependence of the tunneling rate (see Fig. 3).

[46] H. J. Lipkin, N. Meshkov, and A. J. Glick, *Validity of Many-Body Approximation Methods for a Solvable Model: (I). Exact Solutions and Perturbation Theory*, Nucl. Phys. **62**, 188 (1965).

[47] N. Meshkov, A. J. Glick, and H. J. Lipkin, *Validity of Many-Body Approximation Methods for a Solvable Model: (II). Linearization Procedures*, Nucl. Phys. **62**, 199 (1965).

[48] A. J. Glick, H. J. Lipkin, and N. Meshkov, *Validity of Many-Body Approximation Methods for a Solvable Model: (III). Diagram Summations*, Nucl. Phys. **62**, 211 (1965).

[49] M. P. A. Fisher, P. B. Weichman, G. Grinstein, and D. S. Fisher, *Boson Localization and the Superfluid-Insulator Transition*, Phys. Rev. B **40**, 546 (1989).

[50] L. K. Grover, *Quantum Mechanics Helps in Searching for a Needle in a Haystack*, Phys. Rev. Lett. **79**, 325 (1997).